



The Economic Journal, 132 (January), 273–298 https://doi.org/10.1093/ej/ueab062 © 2021 Royal Economic Society. Published by Oxford University Press. All rights reserved. For permissions please contact journals.permissions@oup.com. Advance Access Publication Date: 21 July 2021

# TOWARD AN UNDERSTANDING OF THE ECONOMICS OF APOLOGIES: EVIDENCE FROM A LARGE-SCALE NATURAL FIELD EXPERIMENT\*

#### Basil Halperin, Benjamin Ho, John A. List and Ian Muir

We use a theory of apologies to design a nationwide field experiment involving 1.5 million Uber ridesharing consumers who experienced late rides. Several insights emerge. First, apologies are not a panacea—the efficacy of an apology and whether it may backfire depend on how the apology is made. Second, across treatments, money speaks louder than words—the best form of apology is to include a coupon for a future trip. Third, in some cases sending an apology is worse than sending nothing at all, particularly for repeated apologies and apologies that promise to do better. For firms, *caveat venditor* should be the rule when considering apologies.

'Virtually every commercial transaction has within itself an element of trust .... It can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence.'

Arrow (1972, p. 357)

Economists have come to recognise the importance of trust, reciprocity and other *social preferences* for explaining human behaviour: people are self-interested, but also are often concerned about the payoffs of others (e.g., Rabin, 1993; Charness and Rabin, 2002; Fehr and List, 2004). Additionally, as Arrow (1972) and Sen (1977) have argued, networks of trust and reciprocity are essential for undergirding all economic exchange. However, relatively less is known about the consequences of *violations* of trust or reciprocity. What actions can be taken to avoid the deterioration of mutual confidence when trust has been compromised?

One common action to avoid the collapse of a relationship after a violation of trust or an unfortunate incident is to deliver an apology. The act of apology is an important thread running through households, friendships and employer-employee relationships. Recent research has lent important insights into apologies in lab contexts and small-scale field experiments (see, e.g., Ho, 2012; Gilbert *et al.*, 2017), but much remains ill-understood. For instance, why do firms

\* Corresponding author: Benjamin Ho, Economics, Vassar College, 170 Amsterdam Ave 18E, 10023, New York, USA. Email: beho@vassar.edu

#### This paper was received on 29 January 2020 and accepted on 3 July 2021. The Editor was Rachel Kranton.

The authors were granted an exemption to publish their data because access to the data is restricted. However, the authors provided a simulated or synthetic dataset that allowed the Journal to run their codes. The synthetic/simulated data and codes are available via the Journal website. They were checked for their ability to generate all tables and figures in the paper, however, the synthetic/simulated data are not designed to reproduce the same results. The replication package for this paper is available at the following address: https://doi.org/10.5281/zenodo.5110174.

This manuscript was not subject to prior review by any party, as per the research contract signed at the outset of this project. The views expressed here are solely those of the authors. List was Chief Economist at Uber when the research was completed, and Halperin and Muir were economists on the Ubernomics team. Thanks to Liran Einav and three anonymous reviewers for their insights; seminar participants at AFE 2017, AEA 2018, Williams College and AEA 2019; discussant Chiara Farronato; and Courtney Rosen; for helpful comments and assistance. AEA Registry number: AEARCTR-0002342. All errors are our own.

apologise? Alternatively, if apologies are beneficial, why do firms not apologise *more* often? Do customers actually value apologies?

With these questions as motivating examples, we begin by outlining a principal-agent model of trust violation and apologies. In the model, a customer (the principal) purchases output that provides a noisy signal of the underlying trustworthiness of a firm (the agent). Depending on the stochastic quality of the output, the firm may choose to apologise by sending a (potentially costly) signal to the consumer in an attempt to signal trustworthiness and restore the relationship. Several insights emerge from the model: among them, (1) in order for the apology to be an effective signal, it must be accompanied by a real cost; (2) the efficacy of an apology depends on the familiarity of the consumer with the service; (3) the efficacy of the apology varies by the severity of the outcome; (4) the apology may backfire with repeated use or when an implied promise is broken, i.e., apologising may be worse than not apologising; and (5) apologies may induce substitution from one product line to another within the firm.

We leverage our theory to design a field experiment on the Uber ridesharing platform, which is a natural setting to lend insights into the underpinnings of the model. Uber is concerned that inaccurate estimates of trip duration may lead to decreased trust in the platform and decreased spending in the Uber marketplace. Because Uber services fifteen million rides each day (Bhuiyan, 2018), even a small fraction of rides being late could add up to significant financial repercussions. Indeed, our analysis suggests that, absent any apology, riders who experienced a late trip spend 5%–10% less on the platform relative to a counterfactual rider, suggesting that there are material consequences to the breach of trust described above.

With this substantial loss in revenues as a backdrop, we conduct the first large-scale, natural field experiment to measure the importance of apologies as a method for restoring trust in a relationship. In doing so, we design the experiment to have a tight link with the theoretical model. Our experiment is conducted across the United States over several months, sending real-time apology emails following a late trip, as defined by the actual trip time compared to the initial time estimate shown to the rider. We combine our experimental variation with rich customer data from Uber, the customer-firm relationship history and situational context to test the specific predictions of the model.

A key goal is to measure the role of apologies in maintaining relationships with customers who have received a bad trip experience, measured by the level of future spending with the firm, and then to unpack potential economic mechanisms through which apologies operate. The main set of treatments varies whether a customer receives an apology, the type of apology and the size of the promotional coupon the customer receives as part of that apology (\$5 or zero). We complement these treatments with a secondary set of treatments that send up to two additional apologies following a second and third delayed trip.

We report several interesting insights. First, an apology with a monetary cost after a bad ride—in the form of a \$5 coupon for a future trip—is an effective signal that increases future demand for future trips. Alternatively, we find that an apology with no monetary cost (i.e., words alone) had little effect or was even sometimes counterproductive. As a form of a placebo check, we find that the \$5 coupon administered directly after a bad ride is more effective than a \$5 coupon administered at a random time and unrelated to a rider's experience. We also find that the benefit of an apology with a monetary cost can be detected even three months after the initial bad experience, whereas any benefit from a verbal apology alone quickly fades. This is

especially notable because we measure benefits in this paper as net of the monetary cost of the coupon.

Second, inspired by theory, we consider two other ways verbal apologies create implied economic costs to the firm. These are potential costs incurred by the firm that do not involve a direct monetary payment or coupon. We find that one implied cost is the potential for apologies to backfire, particularly when the apology includes a promise to do better in the future. Our data suggest that in these cases repeated apologies after several bad experiences make things worse for the firm. Firms would have been better off with fewer apologies. Apologies can restore trust but consumers who receive an apology hold firms to a higher standard in the future. If that future expectation is violated, apologies backfire.

Another non-monetary cost implied by a verbal apology that is suggested by theory is a potential cost to the firm's reputation for competence. If the verbal apology acts as an implied admission of incompetence (as suggested by the recent review of Chaudhry and Loewenstein, 2017) then we would expect customers to change the way they use the ridesharing service in the future. In our experimental results, the data lacked the power to make a conclusive statement about this channel.

Finally, we find that characteristics of trips and individuals affect the impact of apologies. The efficacy of an apology depends on the severity of the unsatisfactory service—in this case measured by how late the ride was, in minutes. In particular, we find a U-shaped relationship between severity of the unsatisfactory experience and apology effectiveness: for slightly bad quality and severely poor experiences, apologies are effective. Yet, for moderately poor experiences, apologies are not as effective. Moreover, the efficacy of an apology critically depends on a user's familiarity with the service. Apologies are less effective for users who are quite familiar with the product, yet are much more effective when the user has less experience with the Uber product. Both of these results are in concert with our model.

Our study fits in nicely with several strands of related work. First, it extends the social preference literature into an area that considers how trust can be restored after it is compromised. As Levitt and List (2007) summarise, lab and field experiments with the canonical trust game, dictator game and other games have shown that the concepts of trust and reciprocity are essential for explaining human behaviour. Rabin (1993), Charness and Rabin (2002) and Dufwenberg and Kirchsteiger (2004) formally model these concepts. Second, the extant literature on the economics of apologies has primarily been limited to small-scale field and lab experiments (e.g., Aaker et al., 2004; Abeler et al., 2010; Fischbacher and Utikal, 2013; Chaudhry and Loewenstein, 2017; Gilbert et al., 2017), or difference-in-difference analysis of policy interventions (e.g., Ho and Liu, 2011). We extend this literature by testing the model in the field, with detailed customer and situational data, and we follow the subjects for three months after the apology to measure how effects persist over time. Our data show that methodologically the lab studies have given us a key first look at the efficacy of apologies. Concurrent to our experiment, Cohen et al. (forthcoming) ran a smaller scale apology experiment with the ridesharing platform Via. Our results confirm many of their findings, and extend them by taking advantage of our larger scale and theoretical framework to better understand how and why apologies function.

The remainder of our paper proceeds as follows. We first introduce the principal-agent model that guided the experimental design. Then we provide details of the experimental design, briefly describe the Uber ridesharing platform and discuss the empirical results. We conclude with a

discussion exploring how firms and individuals can use our results to further their understanding of apologies.

# 1. Theoretical Motivation

# 1.1. Defining Apologies

Apologies are hard to define. For example, the Apology Act of 2009, passed in Canada, legally defines an apology as an 'expression of sympathy or regret, a statement that a person is sorry or any other words or actions indicating contrition or commiseration, whether or not the words or actions admit fault or liability or imply an admission of fault or liability in connection with the matter to which the words or actions relate.' Note that an apology does not necessarily require the words 'I am sorry'. It does not necessarily require regret or contrition, but could just be an expression of sympathy. It actually does not even require words at all: according to Canadian law, an apology could be an action. Any attempt to formalise what words or actions constitute an apology are better left to linguists or lawyers. We define apologies based on the economic concepts of reputation and cost, which economists have a comparative advantage in measuring and understanding.

We follow Ho (2012) and define an apology as any costly act that restores a firm's reputation after something has happened that damaged the firm's reputation. The model finds that a cost is necessary to restore the breach in reputation, but that cost does not have to be monetary. When we unpack the mechanism behind apologies, what we are specifically doing is determining the nature of those costs. We acknowledge that there may be features affecting choice beyond the scope of our rational model of apologies. We revisit these possibilities in the Discussion, yet remain focused on the economic costs incurred by the firm who apologises in this study.

When discussing the channels that apologies operate within, we use the term monetary cost apologies to talk about apologies that involve a tangible monetary cost to the firm (in our case a \$5 coupon). We use the term verbal apology to describe the 'cheap-talk' portion of the apology. While the words associated with an apology may indeed be cheap talk, game theory has shown that cheap talk can have real economic consequences (see Online Appendix A), and we argue that words alone can lead to real economic costs to the firm either by causing customers to hold the firm to a higher standard, or by causing the firm to incur a cost to their reputation. The treatments of our experiment were designed to disentangle these different channels.

The second manner in which we deal with the complexity of defining apologies is that we delegate the actual form and phrasing of the apology to the marketing team within Uber. The marketing team was apprised of the type of apology we wanted to convey. They then designed an apology message consistent with their usual corporate standards. In that sense, the apologies we use in our experiment are likely more representative of how a typical corporation might apologise than any message designed by a group of economists, keeping in spirit with our natural field experimental approach.

This does highlight one weakness of our approach. We do not know how exactly the words contained in the verbal apology will be interpreted, and what else the consumer inferred from them. Instead, we did our best to instruct the marketing team on the theory's intent, and attempt to measure the economic consequences of that effort, in terms of consumer behaviour.

t = 0					t = 1
Agent produces	Apology/no apolog	y Principal updates	Outside option	Principal chooses	Old/new agent
$y=y(\theta,\omega)$	$a \in \{0, 1\}$	posterior $b(a, y)$	revealed, $p_{out}$	stay/leave	produces output
	at cost $c(a)$				

Fig. 1. Timeline of the Apology Game.

# 1.2. Setup

Our theoretical framework is based on the Ho (2012) principal-agent model of a customer-firm relationship that formalises many of the findings about apologies in the psychology literature.<sup>1</sup> In the spirit of the Berg *et al.* (1995) trust game, trust is defined as the belief that leads a principal to take an action that relies upon the choices made by an agent. Trust is based on the belief in the trustworthiness of the agent, and our definition of trustworthiness is any characteristic of the agent such as quality or loyalty that leads to higher payoffs for the principal when an act of trust is taken. Trust is a belief in the trustworthiness of the agent. An apology is an action that restores lost trust.

The model is a two-player game between a firm (the agent) and a consumer (the principal). Firms can be a good 'high' type (e.g., high trustworthiness) or bad 'low' type (e.g., low trustworthiness),  $\theta \in \{\theta_H, \theta_L\}$ . The firm produces output y for the consumer, generating utility for the consumer. The quality of the output—how long the ride takes to arrive to the destination relative to expectations in our case—depends on firm type  $\theta$  as well as external circumstance,  $\omega \in \Omega$ , that is uncorrelated with firm type (e.g., unexpected weather). Bad outcomes (i.e., low-quality output) can result from a firm with bad intentions,  $\theta = \theta_L$ , or alternatively from a bad draw from the state of nature  $\omega$ . The consumer is only aware of the overall quality of output  $y = y(\theta, \omega)$ . Type is defined so that the high type yields higher expected output for the consumer.<sup>2</sup>

There may be many dimensions of quality over which a firm may wish to signal their competence. For example, depending on the context and the particular consumer, higher quality could mean better on-time performance, or more responsive customer service, or something else entirely. What all these dimensions have in common is that higher quality represents higher expected future utility for that particular customer. We let  $\theta$  represent any dimension of quality that yields higher expected payoffs for a consumer relative to their outside option.<sup>3</sup>

Within the context of the rideshare industry, the timeline of the baseline game proceeds as follows (Figure 1). The consumer begins with a prior p on the probability that the firm is high type. She then experiences a good or bad outcome for a ride,  $y(\theta, \omega) \in \mathbb{R}$ . Next, the firm chooses to apologise or not,  $a \in \{0, 1\}$ , which has a cost  $c(a) = c(a|\theta, \omega) \ge 0$ . Finally, given the quality of the ride y and apology or non-apology a, the consumer updates her beliefs about the firm's type, learns that an outside option is of high type with probability  $p_{out}$ , and then chooses to stay with the firm or to go with the outside option.

<sup>3</sup> In the context of our rideshare experiment, such outside options include competitors to Uber, public transportation and alternative modes of transport.

© 2022 Royal Economic Society.

277

<sup>&</sup>lt;sup>1</sup> For example, in lab experiments, de Cremer *et al.* (2011) and Ohtsubo *et al.* (2012) find that costly apologies can work better than cheap apologies; Kim *et al.* (2004) and Skarlicki *et al.* (2004) find that apologies can backfire; and many find that the efficacy of an apology depends on the type of offence (e.g., Maddux *et al.*, 2011).

 $<sup>^2</sup>$  To make this example concrete for the rideshare context, a driver may choose an alternate but longer and more costly route than the one the customer initially expected. A driver with 'good intentions' but 'bad circumstances' does so to justifiably avoid heavy traffic or road construction. A driver with 'bad intentions', does so to make more money under the false pretext of traffic or road construction.

The consumer cares only about maximising her utility from rides, where ride quality  $y(\theta, \omega)$  is a function of the firm's type  $\theta$  and external circumstances  $\omega$  such as traffic or weather.<sup>4</sup> The consumer's choice, *x*, is simply whether to purchase from the same rideshare firm in period two or to take an outside option (e.g., switch to a competitor or take public transit):

$$U_{\text{consumer}}(x) = \sum_{t=0,1} y(\theta_t(x), \omega_t).$$

For this application, the firm's problem is simply to decide whether or not to apologise. As in Ho (2012), we abstract away from the firm's choice of effort in the determination of output quality. The firm receives a profit per customer of  $\pi$  and pays the apology cost (potentially zero) given by c(a), which can depend on its type and the state of nature:

$$U_{\rm firm}(a) = \pi + \pi \cdot x - c(a).$$

For the moment, assume that the cost of apologies is constant:  $c(1) = \kappa$ . We discuss other cost functions and cheap apologies (i.e., c(1) = 0) below.

The consumer chooses to stay with the firm provided their posterior belief, given by  $b(a, y) \equiv \Pr[\theta = \theta_H | a, y]$ , is greater than the quality of the outside option:  $b(a, y) > p_{out}$ . The quality of the outside option is drawn from some known distribution  $F(\cdot)$ . The firm chooses to apologise if and only if

$$\pi \cdot [F(b(y, 1)) - F(b(y, 0))] > \kappa.$$

A direct application of Bayes' rule gives us the efficacy of an apology,  $\Delta b \equiv b(y, 1) - b(y, 0)$ , which is the change the apology imparts on the customer's beliefs (i.e., the firm's reputation) and thus the likelihood that the customer will stay with the firm. The model provides several useful predictions about apology efficacy,  $\Delta b$ , that inform our experiment. Below, we discuss how apology efficacy is affected by uncertainty, the costliness of the apology and the severity of the bad outcome. We also discuss predictions regarding repeat apologies.

# 1.3. Equilibrium and Predictions

A separating equilibrium where apologies signal higher type exists given the usual single crossing condition: from Proposition 2 of Ho (2012), there are three existence conditions that allow a separating equilibrium to exist: (1) it is cheaper for high types to apologise, (2) continuing the relationship is more beneficial for high types or (3) high types fail in different situations than low types.

We do not know whether this single crossing will be satisfied. Perhaps the reason we often do not see apologies for a given situation is because single crossing is not satisfied in that situation. For example, it may be cheaper for untrustworthy agents to apologise. The model says that in order to see apologies by rational agents, single crossing is a necessary condition. One factor that helps satisfy single crossing in this case is that single crossing relies on higher types having a higher return from signalling. Since higher quality agents generate more surplus in the future, it is plausible that they also get more value from the relationship.

<sup>&</sup>lt;sup>4</sup> It may seem odd that we refer to type  $\theta$  as intentions when there is no opportunity for the firm to act on those intentions. Instead the model assumes that type exogenously determines output. Ho (2012) showed that, under fairly weak supermodularity conditions, a model where a firm makes a choice that determines the customer outcome, where the choice depends on firm type, produces an isomorphic model to one where firm type directly determines output.

In a separating equilibrium, two properties about the efficacy of an apology follow straightforwardly from Bayes rule (see Ho, 2012 for details).

- 1. Apologies are more effective the greater the apology cost. Furthermore, apologies are *only* effective when there is a cost, c(a = 1) > 0.
- 2. Apologies are more effective when there is greater uncertainty in the relationship (when the prior p is closer to  $\frac{1}{2}$  then to 0 or 1).

We now discuss these two properties in turn, including the intuition behind their proofs and their testable implications.

# 1.3.1. Role of costs

Property 1 says that without a cost, apologies will be ineffective. The intuition is that if apologies increase reputation then all firms will want to apologise, if costs are zero or merely sufficiently low. But if all firms apologies with the same frequency then the efficacy of apologies must be zero. Thus, apologies need to be costly in order to ensure good firms and bad firms apologise at different rates, which creates the separation in beliefs necessary for apologies to function.

#### 1.3.2. Role of uncertainty: relationship length and severity of outcome

Property 2 comes from the fact that, when the prior belief about the firm's type is close to 0 or close to 1, then the posterior belief is unlikely to change much given a single additional signal (the apology) and therefore the apology is likely to be ineffective. Apologies move beliefs the most when the customer is most uncertain.

One immediate implication of property 2 is that apologies are more effective early in relationships. A customer receives more and more signals about a firm's type over time, and as a result, her beliefs converge to either 0 or 1 as time passes. Therefore, apology efficacy is greater early in a relationship. This relationship familiarity could either be with the firm as a whole (how much the customer has used Uber in the past) or with the specific firm product (e.g., how much the customer has used UberPOOL, specifically, in the past).

A second more subtle implication of property 2 is that apologies are less effective for a moderate degree of lateness and more effective for extreme lateness. To see this, consider the distribution of possible outcomes (as measured by minutes late) for a high-type firm  $\theta_H$  versus a low-type firm  $\theta_L$ , as depicted in Figure 2. Here we suppose for illustration that the lateness of a trip is given by a normal distribution, with a lower mean for high-type firms than for low-type firms, and common variance.

In this example, certainty that the firm's intentions are bad is maximised at the mean of the  $\theta_L$  distribution. There is more uncertainty when the ride is less late since the firm is more likely to have had good intentions. Similarly, when the ride is more late, the lateness is more likely to be due to the common shock (e.g., weather or traffic). As a result, we would predict apologies to be least effective for intermediate values of lateness and more effective when barely late or extremely late.

The finding that apologies are more effective when there is greater uncertainty is consistent with lab evidence from Fischbacher and Utikal (2013) and Gilbert *et al.* (2017), who found that apologies are more effective when there is greater uncertainty about the firm's intentions.

© 2022 Royal Economic Society.

279



Fig. 2. Distribution of Lateness of Outcomes (Minutes).

*Notes:* Assume that lateness of outcome is normally distributed, with high-type firms having a lower mean then low-type firms, and common variance (e.g., weather or traffic).

# 1.4. Model Extensions

#### 1.4.1. Repeated apologies

It is also useful to apply the above theory to make predictions regarding the efficacy of repeated apologies. Repeat apologies should be less and less effective as the customer gains experience with the firm. The customer is acquiring more and more information, and therefore is becoming more certain about the firm's type. Therefore, the efficacy of an apology should diminish with increased interaction with the firm. In fact, in an extension to the baseline model described above with screening contracts, Ho (2012) predicted that an apology could even begin to backfire if we assume apologies imply a promise for better behaviour.

A cheap-talk model of repeat apologies predicts a backfire effect if we believe that an apology implies an implicit promise to do better in the future and repeated failure breaks that promise (as seen in the trust game experiment by Schweitzer *et al.*, 2006). A promise that is kept signals higher firm quality while a promise broken is worse than no apology at all.

Imagine the principal (consumer) offers the agent (firm) a menu that says the following: if the firm apologises then the relationship will be continued; however, if the firm is late again, the relationship will be immediately terminated in favour of the outside option. A separating equilibrium exists where good-intention firms apologise and accept the threat of immediate termination while bad-intention firms do not apologise and are judged in the future solely based on their performance (See the Online Appendix of Ho, 2012 for details). In the context of Charness and Dufwenberg (2006) a broken promise signals lack of guilt aversion that serves as a second negative signal about the firm's type.

# 1.4.2. Heterogeneous ride types

A final extension to the baseline model described above is to consider the possibility that the firm cares about its reputation along multiple dimensions of quality. For example, a recent review article by Chaudhry and Loewenstein (2017) argued that people want to be perceived as both warm and competent. But, when a transgression occurs, people do not know if the transgression occurred due to a lack of warmth or a lack of competence. An apology increases the apologiser's

reputation for being warm while decreasing the apologiser's reputation for being competent, while a non-apology does the opposite (see the Online Appendix for more details).

In the case of a ridesharing company, we can think of a reputation for warmth corresponding to better customer service, or more fair pricing, or less deceptive routing. We can think of competence as the ability to navigate around traffic, or the ability to get you to a particular destination in the quickest time possible. One concrete way an apology can harm a firm's reputation is in the case where a consumer may not have noticed that the ride was late. An apology could highlight a mistake that the consumer had never noticed.

So why would a firm ever apologise if the apology just highlights mistakes that were made. We propose that it may make sense to admit one's incompetence if it improves the reputation of the firm overall at the expense of looking incompetent in one particular dimension. This would predict that, while total future rides goes up due to an increase in trust in the company overall, rides similar to the bad one would go down, because the firm has admitted that they are incompetent at providing that type of ride. We look for this change in the composition of future rides when we test for the effects of status cost apologies. We expect the status cost apology to do the most damage to a firm's reputation for competence.

# 1.5. Hypotheses

In sum, the hypotheses from the model that are applicable to our setting are as follows.

HYPOTHESIS 1. The efficacy of an apology is higher when apologies are more costly.

HYPOTHESIS 2. The efficacy of an apology is higher early in the customer-firm relationship, when there is greater uncertainty about the firm's type.

HYPOTHESIS 3. The efficacy of an apology is lowest when trips are moderately late and highest for the most extreme late trips.

HYPOTHESIS 4. The efficacy of an apology decreases with repeated use and can backfire if overused.

HYPOTHESIS 5. An apology decreases future demand for similar trips but increases future demand for dissimilar trips.

The model defines apology efficacy as the change in beliefs  $\Delta b$  that arise in response to an apology. While we do not observe the beliefs of our experimental subjects, we do observe their future decision of whether to stay with the firm, or to choose an outside option:<sup>5</sup> Pr[ $b(\mathcal{H}) > p_{out}$ ]. It is this outcome variable that we use to test our main hypotheses.

# 2. Experimental Design

To test the hypotheses from this model, we conducted a natural field experiment (see Harrison and List, 2004) on the Uber ridesharing platform. The Uber platform connects riders with drivers willing to provide trips at posted rates. A rider provides her desired pickup and dropoff location through a phone app, and is offered a price, an estimated time to pickup and an estimated time to destination (ETD). She then may choose to request an Uber ride and will be picked up and

<sup>&</sup>lt;sup>5</sup> In our experimental setting, we do not have data on available outside options (e.g., competing rideshare services or public transportation).

<sup>© 2022</sup> Royal Economic Society.

transported to the destination. At the end of the trip, the rider has the option to tip the driver.<sup>6</sup> This describes the standard 'UberX' product offering that is the focus of our experiment, but Uber offers products that slightly vary this experience. For example, UberPOOL offers a discounted price but may involve trip detours to pick up multiple riders travelling along a similar route.<sup>7</sup> Our analysis includes only trips using the UberX product.

One measure of platform quality is the accuracy of the ETD provided to riders. Rideshare firms such as Uber are justifiably concerned that inaccurate estimates—'bad rides'—may lead to decreased trust and consequently decreased spending with Uber.<sup>8</sup>

To attenuate the costs of bad trips and to test the power of apologies, we designed a natural field experiment. Our field experiment was conducted over the course of several months in 2017. We selected six of Uber's largest markets to ensure a mix of cities with differing levels of competition between Uber and competing ridesharing platforms, and separately to ensure large enough ridesharing markets to generate a sufficient sample size. The experiment involved 1.5 million subjects across the eight treatment groups described below.

Riders entered the treatment upon experiencing a bad ride, defined as an UberX trip that arrived at the destination n minutes later than the ETD initially displayed to riders when choosing whether to request a trip. The threshold n varied by city based on the city's historical distribution of lateness. The threshold was set so that in expectation only the 5% latest trips would be classified as late in each city, which generally implied a 10–15 minute threshold.

An hour after the end of a bad ride, a customer in a treatment group would receive an email, the content of which varied depending on the treatment group.<sup>9</sup> We then follow all of the customer's future interactions with Uber in the successive twelve weeks.

Following our theory, subjects were divided among eight treatment groups (Figure 3). Half received a one-time \$5 promo code while the other half received no promo code. The promo code conditions were crossed with four different apology types, whose full text is given in Table 1.

- (i) Just \$5 promotion.
- (ii) Basic apology: e.g., 'Oh no! Your trip took longer than we estimated.'
- (iii) Status apology: e.g., 'We know our estimate was off.'
- (*iv*) Commitment apology: e.g., 'We're working hard to give you arrival times that you can count on.'

The wording of each email was designed to follow the spirit of our model and informed by Uber's marketing department. The different kinds of apologies were designed to emphasise different apology mechanisms. In particular, the 'Status cost' apology was designed to amplify the effect of apologies on dissimilar rides (as described in Subsection 1.3.2); and the 'Commitment' apology to emphasise the effect on repeated failures (as described in Subsection 1.3.1). The messages were sent as emails, with subject lines that suggested the nature of the apology and

<sup>&</sup>lt;sup>6</sup> Chandar et al. (2019) and Chandar et al. (2019) studied the economics of tipping on the Uber platform.

<sup>&</sup>lt;sup>7</sup> Cohen *et al.* (2016) also used Uber data to study the demand side of the ridesharing market. A number of other papers use Uber data to examine the supply side; see, e.g., Hall *et al.* (2017) and Cook *et al.* (2021).

<sup>&</sup>lt;sup>8</sup> We completed an analysis using a matching methodology to identify the causal effect on future spending of a rider who experienced a late trip—a 'bad ride'—relative to a statistically identical customer who took an identical ride but which arrived on time. This analysis, which helped to motivate the present study, found that riders in the right tail of the lateness distribution spend 5%–10% less on the platform relative to the counterfactual. These results are discussed in Online Appendix Section C.1.

<sup>&</sup>lt;sup>9</sup> Emails were not sent immediately due to technological constraints.



Fig. 3. *Treatments. Notes:* The experiment was a 4 × 2 design with four apology message types crossed with either a no promo code condition or a \$5 coupon condition.

Name of treatment	Subject line	Email body
Control	N/A	N/A
Basic apology (with or without \$5)	'Oh no! Your trip took longer than we estimated.'	'Your trip took longer than we estimated, and we know that's not ok. We want you to have the best experience possible, and we hate that your latest trip fell short.'
Commitment apology (with or without \$5)	'We can do better.'	'Your trip took longer than expected, and you deserve better. This time we missed the mark, but we're working hard to give you arrival times that you can count on.'
Status apology (with or without \$5)	'We know our estimate was off.'	'We underestimated how long your trip would take—and that's our fault. Every trip should be the best experience possible, and we recognize that your latest trip fell short.'
Just \$5 promotion	'Take \$5 off your next trip.'	'You have places to go and people to see. Enjoy \$5 off your next ride with code GetRiding5.'

Table 1. Email Treatments.

highlighted the \$5 promotion if attached. The theoretical motivations for each apology type are found in Online Appendices A and B.

It is impossible to know exactly how each would be construed by the consumer. We expect there would be overlap between treatments. For example, a basic apology could be interpreted as a promise to do better in the future, while a commitment apology could be seen as an admission of incompetence. The idea though is that the promise to do better is made most salient for the commitment treatment while the admission of incompetence is most salient for the status cost treatment.

Treatment groups were balanced on eight dimensions, with each statistic calculated up until prior to the experiment launch:

- (i) average fare previously faced by a rider,
- (ii) days since signing up with Uber,
- © 2022 Royal Economic Society.

- (iii) lifetime dollars spent on Uber,
- (*iv*) lifetime trip count,
- (v) (number of UberPOOL trips taken)/(number of UberX + UberPOOL trips),
- (vi) number of UberPOOL trips taken,
- (vii) number of UberX trips taken,
- (viii) number of support tickets filed.

Technological limitations meant balancing could only be done for subjects who had signed up for the Uber platform before the start of the experiment. Subjects who joined after the start date were randomly assigned to one of the treatment groups. As a result, because of the large number of subjects, means were significantly different in *t*-tests between groups, but the differences were economically small, as reported in Table 2. Online Appendix B contains further details on experimental design, including the language and imagery contained in the apology email.

In general, we report results for future spending net of any promotions applied ('net spending'), including but not limited to our \$5 promo.<sup>10</sup> For example, if a rider took a single \$8 trip in the seven days following treatment, but used a \$5 promotion on that trip, her level of spending would be reported as \$3. The analysis using gross spending yields similar results. We also consider future trip count, future tipping and the extensive margin of whether the rider took any future trips as outcome variables.

# 3. Results

We begin by presenting the unadjusted means of our main outcome variable, net spending, across the seven treatment groups versus the control group. Figure 4(a) presents average spending by riders over the seven days following the bad ride. The figure can be read as follows: we have 186,584 customers in the control group who had a bad trip. On average, these customers spent (net of promotions) \$45.42 in the seven days after the bad trip. Comparing this to the basic apology group, which had 191,825 subjects, we find that those who received our basic apology spent \$45.86 in the seven days subsequent to a bad trip. This result is significant at the p = 0.045 level using a standard *t*-test of means. We also aggregate the treatments into four categories, shown in Figure 4(b). The categories are: the control group, the treatment group that received<sup>11</sup> just the \$5 promo code ('Just promo'), the three treatment groups that received just an apology email ('Just apology') and the three treatment groups that received both a \$5 promo and an apology ('Promo + apology').

To complement this visualisation of the raw data, we provide Table 3, which reports summary statistics for the full set of outcome variables, again at the seven-day horizon. Note that the differential effect across treatments on trip count and whether a rider takes a future trip is qualitatively the same as the effect on spending in terms of rank order. The effects of different treatments on future tipping behaviour are indistinguishable from zero.

Our main empirical specification regresses the outcome variables of interest for each subject i on the set of eight treatment dummies indexed by j, controlling for the variables  $\vec{X}$  on which

<sup>&</sup>lt;sup>10</sup> Uber at this time sent out a variety of promotions to riders. Other promotions were not directly related to lateness of rides.

<sup>&</sup>lt;sup>11</sup> But did not necessarily use.

	Average fare	Days since signup	Lifetime billings	UberPOOL share	Lifetime trip count	Recent UberPOOL trips	Recent UberX trips	Support tickets
Control	14.339	762.622	1,990.844	0.124	131.44	1.277	5.363	0.896
Basic apology	14.318	757.443	1,973.183	0.124	130.023	1.249	$5.254^{*}$	0.877
Basic apology + promo	14.366	761.054	1,984.287	0.123	131.039	1.27	5.317	0.883
Commitment apology	14.276	759.031	1,963.447	0.124	129.578	1.317	5.289	0.87
Commitment apology + promo	14.383	757.146	1,979.742	0.123	129.618	1.242	5.279	0.863
Status apology	14.309	757.866	1,974.789	0.122	129.4	1.234	$5.222^{***}$	0.87
Status apology + promo	14.356	761.665	1,995.218	0.124	131.619	1.28	5.377	0.893
Just promo	14.368	762.392	1,994.212	0.124	131.528	1.281	5.351	0.886
Notes: * indicates significance of na	irwise t-test versus	the control group at th	ie 5% level. with the B	Sonferroni correction a	unlied ** indicates th	e same at the 1% level an	d *** at the 0.1% leve	-

Table 2. Balance Check—Mean Rider Characteristics by Treatment.

© 2022 Royal Economic Society.

Downloaded from https://academic.oup.com/ej/article/132/641/273/6325164 by MIT Libraries user on 01 November 2022



#### Fig. 4. Mean Spending by Treatment Group.

*Notes:* Panel (a) presents raw mean spending (net of any promotions) by treatment arm. Panel (b) aggregates the results across message types into four treatment categories (since the content of the messages themselves was found to be insignificant), with shaded 95% confidence intervals. Sample sizes are rounded to the thousands for ease of reading. Note that the commitment apology arms have smaller samples because of the way that the 'repeated apologies' sub-experiment was structured; see Subsection 3.3.

	Total spending (net of promos)	Trip count	Total spending (incl. promos)	Total tips	Took another trip
Control	45.424	2.848	46.6	0.977	0.674
	(0.152)	(0.009)	(0.154)	(0.008)	(0.001)
Just promo	45.741	2.877	47.219	0.994	0.680
	(0.154)	(0.009)	(0.156)	(0.009)	(0.001)
Just apology	45.748	2.851	46.924	0.991	0.672
	(0.100)	(0.006)	(0.101)	(0.006)	(0.001)
Promo + apology	45.649	2.879	47.166	0.994	0.679
	(0.101)	(0.006)	(0.102)	(0.006)	(0.001)

Table 3.	Means	(SEs) by	Treatment	Category.	Seven-Day	Horizon.
rable 5.	means	(DLS) Uy	mannen	Curegory,	Seven Duy	110112,011.

Notes: Outcome variables at a seven-day horizon are presented here, but data were collected at horizons up to and beyond eighty-four days after the initial bad ride.

we balanced in addition to city, date, and hour-of-week fixed effects:

$$\log(\text{Outcome}_i) = \sum_j \alpha_j \cdot \text{Treatment}_j + \vec{\beta} \cdot \vec{\mathbf{X}}_i + \gamma_{\text{city}} + \delta_{\text{date}} + \eta_{\text{hour}} + \varepsilon_i. \tag{1}$$

Regression results for the effect of apologies on net spending, estimated using this specification, are presented in Table 4.<sup>12</sup> Each column estimates the treatment effect on net spending over progressively longer horizons (7, 14, 28, 56, 84 days).

<sup>12</sup> Online Appendix Table D.1 presents the results of the same regression using indicator dummies for the grouped treatments described previously.

	7d	14d	28d	56d	84d
Basic apology	0.007	0.001	-0.005	-0.008	-0.009
1 00	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
Basic apology $+$ promo	0.015**	0.012*	0.015**	0.011*	0.009
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
Commitment apology	-0.002	-0.004	-0.013	$-0.016^{*}$	-0.014
	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)
Commitment apology + promo	0.008	-0.001	0.001	-0.005	-0.006
	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)
Status apology	0.006	0.002	-0.003	-0.006	-0.006
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
Status apology + promo	0.013*	0.008	0.011	0.011*	0.011*
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
Just promo	0.015**	0.007	0.005	0.005	0.008
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
Controls	Х	Х	Х	Х	Х
City	Х	Х	Х	Х	Х
Date	Х	Х	Х	Х	Х
Hour	Х	Х	Х	Х	Х
No. observations	1,257,738	1257,737	1,257,735	1,257,735	1,257,740

Table 4. Log of Future Net Spend by Treatment Group over the N days after the Bad Ride.

*Notes:* OLS regressions of log future net spending in the *N* days after experiencing a bad ride with city, date and hourof-week fixed effects. Controls include: average fare; days since signup; lifetime billings; lifetime POOL share; lifetime trips; number of recent POOL trips; number of recent UberX trips and number of support tickets filed. \*\*\* p < 0.001, \*\* p < 0.01, \*p < 0.05.

A main feature to note is that the verbal apology by itself (without a promotion) has no statistically significant effect at conventional levels. In fact, while the effect of a verbal apology is largely not significant, if anything the presence of the verbal apology in and of itself has a negative point estimate over longer time horizons (fifty-six to eighty-four days). Table 5 presents the same specification but with number of future rides as an outcome variable. It shows the same basic pattern, and we therefore focus our attention on net spending as the outcome variable.

Figure 5 plots the estimated coefficients on the treatment dummies from our main empirical specification estimated over the same horizons described above. We find persistent effects of treatments that include a promotion—which, recall, was a one-time promotion—as far out as three months after the apology was sent.

One possible explanation for the persistence of the effect is intertemporal complementarities in consumption. In other words, if taking an additional ride today increases a rider's chance of taking a ride tomorrow then simply inducing a customer to take an additional ride in the first week could have persistent effects. While this result is intuitively appealing, it should be tempered in that if complementarities were the only force driving the persistence, one would expect the effect size to get smaller over time. In fact, if anything the effect (of a promotional coupon alone) stays steady or increases (albeit not significantly) by day eighty-four.

What is especially notable in the results is that the effect of a verbal apology by itself becomes more negative over time. A verbal apology alone (with no coupon) becomes significantly negative by day eighty-four, in contrast to the effect of an apology with a promo. In particular, the difference in effects of an apology without a coupon by day seven is statistically distinguishable from the effect by day eighty-four ( $p = 3.6 \times 10^{-5}$ ), whereas the difference for the effect of an apology including a coupon is not (p = 0.33). These points confirm

© 2022 Royal Economic Society.

287

		nue.			
	7d	14d	28d	56d	84d
Basic apology	0.004	0.001	-0.003	-0.005	-0.006
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Basic apology + promo	0.011***	0.01***	0.01***	0.008*	0.007*
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Commitment apology	-0.001	-0.002	$-0.008^{*}$	$-0.011^{*}$	$-0.011^{*}$
	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)
Commitment apology + promo	$0.008^{*}$	0.003	0.002	-0.001	-0.002
	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)
Status apology	0.003	$2.68 \times 10^{-04}$	-0.003	-0.006	$-0.007^{*}$
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Status apology + promo	0.01***	0.008**	0.01***	0.01**	0.01**
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Just promo	0.01***	$0.007^{*}$	$0.007^{*}$	$0.007^{*}$	$0.008^{*}$
-	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Controls	Х	Х	Х	Х	Х
City	Х	Х	Х	Х	Х
Date	Х	Х	Х	Х	Х
Hour	Х	Х	Х	Х	Х
No. observations	1,257,788	1,257,788	1,257,788	1,257,788	1,257,788

 Table 5. Log of Future Number of Rides by Treatment Group over the N days After the Bad

 Ride

*Notes:* OLS regressions of log future trips taken in the *N* days after experiencing a bad ride with city, date and hourof-week fixed effects. Controls include: average fare; days since signup; lifetime billings; lifetime POOL share; lifetime trips; number of recent POOL trips; number of recent UberX trips and number of support tickets filed. \*\*\* p < 0.001, \*\* p < 0.01, \*p < 0.05.

Hypothesis 1 that apologies are more effective when the cost associated with the apology is higher.

Breaking the results down by treatment, we can see in Figure 5(a) that the downward time trend is seen primarily in the treatments with no coupon, along with both coupon and no-coupon treatments when a commitment was made.<sup>13</sup> There were declines in some of the other treatments, but they were smaller and not statistically significant.

One possible explanation for the smaller declines in the non-commitment treatments is that theory predicts that backfire occurs when a promise has been broken. The commitment treatment was designed to increase the salience of the implicit promise that comes with apologies, and that is where we saw the greatest backfire effect, but it is likely that a promise to do better was inferred by consumers in the other treatments as well, just to a smaller extent.

We also have data on who decided to actually open the apology email. Although we do not know which customers saw the subject line or snippet but chose not to open it, we can think of those who actually opened the email as the treated group, and do a local average treatment effect analysis using the subject's treatment group assignment as the instrumental variable. Approximately one-third of people who received the email opened it, although the exact number varied somewhat by treatment.<sup>14</sup> Online Appendix C reproduces our main results and figures using this instrumental

 $<sup>^{13}</sup>$  The one treatment in this set that did not see a statistically significant decline at a 5% significance level was the commitment apology without a promo. However, the estimated coefficients for this treatment had a decline similar to the others, around 1.2%, with a *p*-value of 0.098. The lower significance for commitment apologies could be explained by the lower power in that treatment as we sub-divided the commitment treatment into eight groups to measure repeat apologies, so the sample we are testing is only one-eighth as large as the other treatments.

<sup>&</sup>lt;sup>14</sup> In particular, the 'commitment' treatment was 10.28 percentage points less likely to be opened than the 'basic treatment' (p < 0.001); and the 'status' treatment was 0.83 percentage points less likely to be opened than the 'basic' treatment (p < 0.001).







*Notes:* We plot the  $\alpha$  coefficient on each treatment dummy from model (1), with total spending as the outcome, between the date of the bad ride and some future date 7, 14, 28, 56 and 84 days in the future. Panel (a) presents results for each treatment, and panel (b) aggregates the results across message types into four treatment categories for increased power. We include 95% confidence intervals in panel (b).

variables approach. We find the same patterns of magnitude and significance, but the coefficients are approximately three times larger. For example, a basic apology with a promo led to a 2.7% increase in net revenues over eighty-four days, while a commitment apology made without a promo led to 6.5% decline. This suggests that our effects are driven almost entirely by those who opened the email.



Fig. 6. Rider Heterogeneity.

*Notes:* Treatment effect on net spending within sub-samples defined by the quartile of the number of past rides in the customer's history.

## 3.1. Heterogeneity by Rider History

We now turn to Hypothesis 2 that apologies should be most effective when there is the greatest degree of uncertainty and therefore we would expect greater efficacy for new users of the ridesharing platform. Here we present the effect of apologies within sub-samples of riders based on quartiles of riders' numbers of rides before having the bad experience.

As shown in Figure 6, our results are quite mixed. For the joint promo and apology treatment, the point estimates indicate that the treatment effect is highest for the newest quartile of users (those with 0 to 10 lifetime trips) and lowest for the most experienced users (those with greater than 157 trips), with the effectiveness decreasing across quartiles.<sup>15</sup> However, for those who received just a verbal apology, the point estimates are mixed, and in fact the treatment effect estimate is highest for the most experienced users and somewhat lower for the newest users. In all cases, the confidence intervals are wide and statistically indistinguishable.

Looking instead at a different measure of unfamiliarity and uncertainty, the frequency of UberPOOL usage relative to UberX, we find results directionally consistent with the hypothesis but again statistically indistinguishable (Figure 7). The two most popular services provided by Uber are UberPOOL and UberX. Since our experiment was conducted exclusively on UberX riders, we expect riders who have mostly used UberPOOL in the past to be more uncertain about the quality of UberX. Indeed, our point estimates indicate that riders who mostly used UberPOOL in the past were much more likely to be positively influenced by an apology than riders who mostly used UberX, although the confidence intervals are again large.

<sup>&</sup>lt;sup>15</sup> An alternate explanation for this finding is that experienced users have received a large number of promos by the time they have completed 157 trips, and may perceive the promo as a 'low cost' apology, whereas new users, having received fewer promos in the past, value the promo more and perceive them to be a 'higher cost' apology.



# Percent change in net spend (7d)

Fig. 7. UberPOOL Riders versus UberX Riders.

*Notes:* Treatment effect on net spending within sub-samples defined by whether a rider is a 'consistent UberX user' ( $\geq$  75% of trips in the preceding three months on UberX), a 'consistent UberPOOL user' ( $\geq$  75% of trips in the preceding three months on UberPOOL) or the intermediate case with no consistent product.

# 3.2. Heterogeneity by Severity of Lateness

Recall Hypothesis 3 that an apology would be least effective for moderate levels of lateness, since this is when the poor experience is most likely attributable to the firm itself. On the other hand, apologies would be more effective for low levels of lateness (when the firm is more likely to be of the high type) and high levels of lateness (where the most severe delays can be attributed to external factors like weather).

We test this relationship formally by estimating our main specification (1) with the addition of interaction terms between the treatment dummies and the percentile of lateness and the percentile squared.<sup>16</sup> Since there is significant variation in the distribution of lateness for each city, we measure lateness relative to other rides from the same city, although other specifications produce the same pattern.

As shown in Table 6 and visually in Online Appendix Figure D.1, we find that the data are consistent with this prediction: the quadratic interaction term is positive and statistically significant for the 'Promo + apology' treatment at the p = 0.094 level (and for the 'Just promo' treatment at the p = 0.024 level). As predicted, apologies are least effective (or most damaging) for intermediate degrees of lateness. This is also suggestive that the 'Just promo' treatment may

<sup>16</sup> Formally,

$$log(Outcome_i) = \sum_{j} (\alpha_j \cdot \text{Treatment}_j) + (\gamma_1 \cdot \text{percentile}) + (\gamma_2 \cdot \text{percentile}^2) + \sum_{j} (\delta_j \cdot \text{Treatment}_j \cdot \text{percentile}) + \sum_{j} (\rho_j \cdot \text{Treatment}_j \cdot \text{percentile}^2) + \vec{\beta} \cdot \vec{X}_i + \gamma_{\text{city}} + \delta_{\text{date}} + \eta_{\text{hour}} + \varepsilon_i.$$

The coefficient of interest is  $\rho_j$ , the coefficient on the interaction between the treatment dummy and the quadratic term.

© 2022 Royal Economic Society.

291

	7d
Just apology	0.023
1 00	(0.015)
Just promo	0.034
	(0.018)
Promo + apology	0.031*
	(0.015)
Percentile	0.001
	(0.001)
Percentile squared	$-1.24 \times 10^{-05}$
	$(5.78 \times 10^{-06})$
Just apology: percentile	-0.001
T / /1	(0.001)
Just promo: percentile	-0.002
Droma Lanalogy: paraantila	(0.001)
Fiolito + apology. percentite	-0.001
Just apology: percentile squared	(0.001)
Just apology. percentile squared	$(6.88 \times 10^{-06})$
Just promo: percentile squared	$(0.08 \times 10^{-0.05})$
Just promo, percentile squared	$1.65 \times 10^{-06}$
Droma Landory noroantile sourced	$(8.10 \times 10^{-05})$
Promo + apology: percentile squared	$1.10 \times 10^{-06}$
Controls	$(0.89 \times 10^{-0.0})$
City	X
Date	X
Hour	X
No. observations	1,203,471

Table 6. Treatment Effect by Lateness Percentil	e and	Percer	ıtile
Squared.			

*Notes:* \*\*\* p < 0.001, \*\* p < 0.01, \*p < 0.05.

have partially functioned as a type of apology even if we never included any verbal apology to go along with it (see also the Discussion section below).

# 3.3. Impact of Repeat Apologies

Next, consider Hypothesis 4 that repeat apologies would be less and less effective over time and may even be counterproductive. For a sub-sample of riders, we conduct the following secondary experiment. We split the sample and offer a second apology for half of the subjects who in following weeks receive a second late trip, leaving the other half as a control (having only received one apology). For the sub-sample that received two apologies, we split the sample again for those who took a third late trip, offering half a third apology and leaving the other half as one final control (who only received two apologies). Online Appendix Figure D.4 shows this experimental design diagrammatically.

We present the marginal effects of the second and third apology in Figure 8. As before, a verbal apology alone without the \$5 promotion remains largely ineffective. However, whereas the short-term effect of the first apology with a \$5 promotion yielded a 2% increase in net spending, the net effect on spending of the second apology is not significantly different compared to someone who had a second bad ride but received no new apology message. For the third bad ride, the apology on its own is insignificant again, while the third apology with a promotion has





*Notes:* Panel (a) reports the marginal treatment effect *over the first apology* of a second apology treatment for a second bad experience with Uber, compared to the relevant control group. Panel (b) reports the marginal effect of a third apology relative to the second. Effects are at a seven-day horizon.

a significantly negative effect on future net spending relative to someone who received three bad rides but only received two apologies with a promotion. In fact, this negative effect shows up not just in terms of future net spending but also in terms of the number of future rides taken and in terms of future gross spending, and persists over time.<sup>17</sup>

This backfire effect we observe is consistent with a verbal apology that is acting as a promise. A promise-based commitment apology can temporarily restore a customer's loyalty after an adverse outcome, as described in the model of Subsection 1.3.1. However, a commitment apology acts as a promise that the adverse outcome was due to unexpected external factors, and that the customer should therefore expect better outcomes in the future. When those higher expectations go unmet, the firm suffers more than if no apology had been tendered at all. Apologies should therefore be used sparingly and ideally only after unexpectedly bad outcomes that are unlikely to repeat again in the near future.

# 3.4. Heterogeneity by Ride Type

Finally, consider Hypothesis 5 that an apology after a trip of a given type decreases demand for that category of ride but increases demand for dissimilar trips. Given the rich data associated with each trip on the Uber platform, we are able to classify trips into several natural categories based on popular Uber use cases. We consider the following categories of trips: rides to and from an airport, rides during rush hour and rides during weekend hours. We also link trip timing and location to local weather conditions using the Dark Sky weather API and consider trips during times of precipitation (i.e., rain, snow or sleet) versus those not during times of precipitation.

To test the model's hypothesis about heterogeneous effects due to the circumstance of the bad ride, we compared the treatment effect of apologies on riders who had (for example) a bad airport trip on future airport trips versus the treatment effect on future non-airport trips, However, we

293

<sup>&</sup>lt;sup>17</sup> Online Appendix Figures D.2 and D.3 show these treatment effects at different horizons.

<sup>© 2022</sup> Royal Economic Society.

are unable to reject the null that apologising has no differential effect between trip types for the categories tested (airport versus non-airport, rush hour versus non-rush hour, weekend versus weekday and rainy versus non-rainy). These analyses are included in Online Appendix C.5.

# 4. Discussion

Since our principal finding is that it is primarily a promotional coupon that can be used for a future trip that restores the firm's reputation and not the apology itself, one can ask: is this an 'apology effect' or just a 'promo effect'? Perhaps our effects are driven by the fact that a promo code induces customers to take more and longer rides that they would not have otherwise. One approach to answer this query is to compare our estimated effect sizes with the effect of a generic \$5 promotion sent out randomly by Uber, which will have no apology connotation.

Running concurrently in the cities where our experiment was conducted (between the months of June and October of 2017), another experiment tested the effects of randomly sending a \$5 promo against a control group that received no promotion. While this serves as an important comparison experiment, we should note that this natural field experiment is not a perfect analogue to our main apology experiment for two reasons. First, this experiment proactively targeted the entire Uber rider population whereas our own experiment targeted only those who had received a late ride. Having a late ride is more likely to happen to more frequent riders simply by chance: more trips implies a higher chance of at least one bad draw. To make treatment effects comparable, we restrict consideration to just those riders who experienced at least one bad ride during 2017. It is important to note that while these riders experienced a bad ride, the random \$5 promos were not sent because of this ride and could have been sent months before (or after) the experience.

A second limitation of our comparison experiment is that this generic promo was usable multiple times and limited to a single week, whereas our promotion was one-time use in the next three months. Therefore, we might expect this generic promotion to be much more effective at the seven-day horizon than our apology promotion.

In fact, while the sample size is small (n = 27, 203), we find that our 'just promotion' treatment in the aftermath of a bad experience is statistically significantly more effective than the randomly timed generic promotion. Stacking the generic promotion data with our 'just promotion' and control data, we estimate

$$log(Outcome_i) = \alpha_1 \cdot is\_generic + \alpha_2 \cdot is\_treated + \alpha_3 \cdot is\_generic \cdot is\_treated + \vec{\beta} \cdot \vec{X}_i + \gamma_{city} + \delta_{date} + \varepsilon_i,$$
(2)

where the coefficient on the interaction term  $\alpha_3$  is the treatment effect of receiving a generic (randomly timed) promotion, compared to receiving a promotion in the aftermath of a late trip.

Estimating (2) with net spend as the outcome variable at the seven-day horizon, and using the same set of controls as in the previous analyses, we find that the randomly timed promotion has a significantly negative effect of  $\alpha_3 = -8.3\%$  (*p*-value < 0.001) on future net spending versus a promotion after a late ride. Importantly, this suggests that it matters that the act of remediation occurred after an adverse event, a breach of trust. This is, at least, consonant with the idea that our \$5 'just promotion' treatment had an extra impact after a bad trip compared to the effect observed after a generic \$5 promotion is received.

Another possible concern regarding our promo structure is that customers were given three months to use the promo. Therefore, the longevity of our effect, the fact that we saw a consistent one percent increase in future net revenues across the twelve weeks of our data collection, could be due to the fact that customers were saving their promo codes and using them three months later. We acknowledge this is a possibility that we are not able to test for. However, we think that this is unlikely for three reasons. (1) Past experience with other promotional offers finds that most promo codes are much more likely to be used sooner rather than later. (2) Customers in our sample average one ride every six days. Twelve weeks is a long time to wait to use a promo code. (3) A similar study conducted by another ridesharing company that we discuss next, automatically applied a \$5 credit, and also found a persistent effect for the four weeks they tracked their customers.

In general, our findings line up with the findings of an experiment that coincidentally was run independently, and concurrently, by the ridesharing company Via; see Cohen *et al.* (forthcoming). This study also found that, while a \$5 promo after a bad ride was effective at increasing net spending, a \$5 promo randomly given had an insignificant effect on gross and net spending. Cohen *et al.* (forthcoming) also found that a cheap-talk apology (without a promotional coupon) had no significant effect. This replication with a different company is encouraging in that it suggests our results generalise to rideshare firms beyond Uber, which had perhaps a unique reputation at the time our experiment was conducted. There are a couple differences observed between the Cohen *et al.* (forthcoming) paper and our own that are worth noting. They found that apologies mostly matter for late pickups, whereas our experiment focused on late arrivals. Indeed, they found null results for late arrivals. They also found that their apologies are most effective for their most frequent customers whereas we find indistinguishable treatment effects on users by frequency. These differences are likely due to Via's model that emphasises shared rides. When a user hails a ride with Via, she knows that the driver will pick up other riders along the way. Thus, she does not necessarily have the same expectation for an on-time arrival.

The Via study, occurring in a different geography and different setting, is also informative because apologies are undoubtedly context dependent. Abeler *et al.* (2010), who studied apologies on an auction website, is similarly complementary. Interestingly, they found that cheap apologies were more effective than monetary compensation. We have two possible explanations for the incogruence between our results and the insights of Abeler *et al.* (2010). First, their outcome variable was the customer's rating of the seller on the auction website. This is relatively costless for the customer to change. The second is that their offer of monetary compensation was offered as a quid pro quo payment to the customer to change the rating (which may have been construed as a bribe), whereas in our case the monetary compensation was offered as a gift. Of course, our thoughts are merely speculative, and further experiments are needed to precisely identify the role of norms and context. Both Cohen *et al.* (forthcoming) and Abeler *et al.* (2010), along with the present study, are limited to analysis of apologies by firms, rather than by individuals. This may matter: for instance, if it is easier for an individual to admit incompetence compared to a multi-agent firm; or commitments to improve the future are more credible coming from a firm trying to protect its reputation in repeated interactions.

A comparison to the Via study also highlights one other advantage of a large dataset. Even with 1.5 million people in our subject pool, many of our heterogeneity tests were still underpowered. The size of our sample was what made possible any analysis of heterogeneity at all.

Future research can also better identify the mechanisms that determine how apologies work. Apologies can contain monetary restitution, admission of guilt, promises about the future, expression of empathy or even excuses (see Online Appendix A for more details). The experiment was designed to test different apology mechanisms by varying the message that accompanied

© 2022 Royal Economic Society.

the apology and by estimating the effect of apologies in different traffic and weather situations. While some of the effects of different apology messages were consistent with predictions from theory (e.g., promise apologies were most likely to backfire), the significance of other effects were not consistently robust, perhaps because the email text associated with promotions was not carefully read by customers (email open rates averaged approximately 30%).<sup>18</sup>

Similarly, the efficacy of apologies did vary by weather and traffic, but not in a way that was statistically significantly discernible by our theory. One hope of testing for the effects of weather and traffic was to see if the fault of the lateness made a difference, but such a pattern was not evident in our data. We also tested the effect of status apologies by examining if increased demand for one type of ride was offset by decreased demand for other types of ride. The model was inspired by findings that an apology often increases the apologiser's perceived warmth at the expense of their perceived competence. However, our tests of such trade-offs turned up no significant results. Perhaps unsurprisingly, applying an apology concept designed to convey warmth may not be a good fit when it comes to corporate apologies.

It is also worth taking a moment to consider the time span of our results. We found significant negative and positive effects of apologies at the end of the eighty-four days of observation. In the results, we focused on the backfire effects seen for the commitment apologies, but we also saw apparently smaller backfire effects in the 'cheap' apologies in the last column of Table 4. The significance of these differences are not robust, but smaller backfire effects are consistent with theory for types of apologies where the potential for gain is also small.

One might also wonder why apologies continue to impact consumer behaviour eighty-four days later. One distinct possibility is that the apology changes the consumer's perception of the firm and that perception affects how consumers view future encounters with the company. Another possibility is that the apology increases the consumer's likelihood to take an Uber ride in the following week or two, which creates more opportunities for Uber to demonstrate their trustworthiness as a company. We leave that distinction for future work.

Finally, we note that our investigation is driven by the assumptions of our costly signalling model, but of course there could be other psychological mechanisms at work that we have not considered. For example, consumers may have interpreted the actions of Uber with bemusement, or reciprocity, or contempt. One could think of those other psychological mechanisms in the 'as-if' framework articulated by Frank (1988), where such mechanisms operate by having us behave 'as-if' agents were playing the rational strategy in a costly signalling game. While we did not follow that path, a fuller account of what consumers are thinking when they receive an apology is a fruitful area for future research.

# 5. Conclusion

We present results from a large-scale natural field experiment on the effects of apologies to restore trust within a principal-agent relationship. We offer not just evidence that apologies matter for customers, but also insights into how apologies matter. Our results have implications both for firms deciding how, and when, to apologise and for understanding how trust can be repaired in economic relationships more generally.

The main finding is that apologies with a monetary cost was the primary channel through which apologies repair relationships. Verbal apologies alone without a monetary cost were

<sup>&</sup>lt;sup>18</sup> For additional analysis of the local average treatment effect of *opening* emails, see Online Appendix C.2.

largely ineffective. However, words do matter. Verbal apologies that contained a promise to do better in the future led to worse outcomes for the firm compared to the case of just sending the \$5 promotion.

Specifically, we find that the most effective apology was the provision of a \$5 coupon, with or without any accompanying apology text. Giving such a coupon after a bad ride was more cost effective than \$5 coupons given at random. This provides direct evidence for Hypothesis 1, that more costly apologies are more effective. We further examine dimensions of customer characteristics and characteristics of the adverse outcome that could help provide guidance for more effective apologies going forward, such as the customer's familiarity with the product, though we do not find supporting evidence for the prediction of Hypothesis 2 that apologies are more effective for new customers. In the non-linear relationship between lateness and apology efficacy, we find direct evidence for Hypothesis 3, apologising with a promise to do better or apologising repeatedly to the same person who had multiple bad experiences actually reduced future spending in the short run, relative to someone who also had similar bad rides but did not receive an apology, as predicted by Hypothesis 4. Finally, we are unable to detect supporting evidence for Hypothesis 5, that an apology decreases demand for similar trips but increases demand for dissimilar trips.

Overall, our experiment provides a degree of real-world empirical support for the general apology model. While previous lab studies have served to provide important insights, our data demonstrate the value of the signalling view of apologies by showing that some of its predictions hold in the field: apologies with more monetary cost are more effective than apologies without monetary cost; the efficacy of the apology varies by the severity of the outcome; an apology may backfire with repeated use or when an implied promise is broken.

Our analysis also provides useful advice for firms on the ifs, whens, wheres and hows to apologise optimally. We find that, while apologies can be an effective way to restore and prolong the customer relationship, the reason why apologies are not more frequent is because they are costly and potentially backfire. Firms often do not apologise because apologising is difficult. Our data highlight that the safest way to remediate a bad experience is a simple promotion applied to future purchases. We find that money spent in this way, after an adverse event, yields a positive return for the firm even when promotions sent at other times do not.

There are several opportunities to expand on our experiment. Future work should explore the impact of apologies in other industries and include greater variation in the cost dimension. In particular, we remain interested in exploring the role of different kinds of apologies where the implicit promises associated with an apology are made more explicit. The wording of the messages sent in our experiment were restricted by the constraints imposed by Uber's marketing practices, and limitations in our ability to ask customers what they thought about them. We were hoping to convey a commitment to do better, or an admission of incompetence. Future work could make those messages more precise.

Massachusetts Institute of Technology, USA Vassar College, USA University of Chicago & NBER, USA Lyft, Inc., USA Additional Supporting Information may be found in the online version of this article:

# Online Appendix Replication Package

# References

- Aaker, J., Fournier, S. and Brasel, S.A. (2004). 'When good brands do bad', Journal of Consumer Research, vol. 31(1), pp. 1–16.
- Abeler, J., Calaki, J., Andree, K. and Basek, C. (2010). 'The power of apology', *Economics Letters*, vol. 107(2), pp. 233-35.
- Arrow, K.J. (1972). 'Gifts and exchanges', Philosophy & Public Affairs, vol. 1(4), pp. 343-62.
- Berg, J., Dickhaut, J. and McCabe, K. (1995). 'Trust, reciprocity, and social history', *Games and Economic Behavior*, vol. 10(1), pp. 122–42.
- Bhuiyan, J. (2018). 'Uber powered four billion rides in 2017. It wants to do more—and cheaper—in 2018', Recode.
- Chandar, B., Muir, I., List, J.A. and Gneezy, U. (2019). 'The determinants of tipping: evidence from rideshare', Working Paper, National Bureau of Economic Research.
- Chandar, B., Muir, I., List, J.A., Woolridge, J.M. and Hortaçsu, A. (2019). 'Design and analysis of cluster-randomized field experiments in panel data settings', Working Paper, National Bureau of Economic Research.

Charness, G. and Dufwenberg, M. (2006). 'Promises and partnership', Econometrica, vol. 74(6), pp. 1579-601.

- Charness, G. and Rabin, M. (2002). 'Understanding social preferences with simple tests', *The Quarterly Journal of Economics*, vol. 117(3), pp. 817–69.
- Chaudhry, S. and Loewenstein, G. (2017). 'Thanking, apologizing, bragging, and blaming: the currency of communication', *Psychological Review*, vol. 126(3), pp. 313–44.
- Cohen, M.C., Fiszer, M.D. and Kim, B.J. (forthcoming). 'Frustration-based promotions: field experiments in ride-sharing', Managment Science, pp. 1–21.
- Cohen, P., Hahn, R., Hall, J., Levitt, S. and Metcalfe, R. (2016). 'Using big data to estimate consumer surplus: the case of Uber', Working Paper, National Bureau of Economic Research.
- Cook, C., Diamond, R., Hall, J., List, J.A. and Oyer, P. (2021). 'The gender earnings gap in the gig economy: evidence from over a million rideshare drivers', *The Review of Economic Studies*, vol. 88(5), pp. 2210–38.
- de Cremer, D., Pillutla, M.M. and Folmer, C.R. (2011). 'How important is an apology to you? Forecasting errors in evaluating the value of apologies', *Psychological Science*, vol. 22(1), pp. 45–8.
- Dufwenberg, M. and Kirchsteiger, G. (2004). 'A theory of sequential reciprocity', *Games and Economic Behavior*, vol. 47(2), pp. 268–98.
- Fehr, E. and List, J.A. (2004). 'The hidden costs and returns of incentives-trust and trustworthiness among CEOs', Journal of the European Economic Association, vol. 2(5), pp. 743–71.
- Fischbacher, U. and Utikal, V. (2013). 'On the acceptance of apologies', *Games and Economic Behavior*, vol. 82(C), pp. 592–608.
- Frank, R.H. (1988). Passions Within Reason: The Strategic Role of the Emotions, New York: W. W. Norton & Company.
- Gilbert, B., James, A. and Shogren, J. (2017). 'Corporate apology for environmental damage', *Journal of Risk and Uncertainty*, vol. 56(1), pp. 51–81.
- Hall, J.V., Horton, J.J. and Knoepfle, D.T. (2017). 'Labor market equilibration: evidence from Uber', Working Paper.
- Harrison, G.W. and List, J.A. (2004). 'Field experiments', Journal of Economic Literature, vol. 42(4), pp. 1009-55.
- Ho, B. (2012). 'Apologies as signals: with evidence from a trust game', Management Science, vol. 58(1), pp. 141-58.
- Ho, B. and Liu, E. (2011). 'Does sorry work? The impact of apology laws on medical malpractice', *Journal of Risk and Uncertainty*, vol. 43(2), pp. 141–67.
- Kim, P.H., Ferrin, D.L., Cooper, C.D. and Dirks, K.T. (2004). 'Removing the shadow of suspicion: the effects of apology versus denial for repairing competence- versus integrity-based trust violations', *Journal of Applied Psychology*, vol. 89(1), pp. 104–18.
- Levitt, S.D. and List, J.A. (2007). 'What do laboratory experiments measuring social preferences reveal about the real world?', *Journal of Economic Perspectives*, vol. 21(2), pp. 153–74.
- Maddux, W.W., Kim, P.H., Okumura, T. and Brett, J.M. (2011). 'Cultural differences in the function and meaning of apologies', *International Negotiation*, vol. 16(3), pp. 405–25.
- Ohtsubo, Y., Watanabe, E., Kim, J., Kulas, J.T., Muluk, H., Nazar, G., Wang, F. and Zhang, J. (2012). 'Are costly apologies universally perceived as being sincere?', *Journal of Evolutionary Psychology*, vol. 10(4), pp. 187–204.
- Rabin, M. (1993). 'Incorporating fairness into game theory and economics', *The American Economic Review*, vol. 83(5), pp. 1281–302.
- Schweitzer, M.E., Hershey, J.C. and Bradlow, E.T. (2006). 'Promises and lies: restoring violated trust', Organizational Behavior and Human Decision Processes, vol. 101(1), pp. 1–19.
- Sen, A. (1977). 'Rational fools: a critique of the behavioral foundations of economic theory', *Philosophy & Public Affairs*, vol. 6(4), pp. 317–44.
- Skarlicki, D.P., Folger, R. and Gee, J. (2004). 'When social accounts backfire: the exacerbating effects of a polite message or an apology on reactions to an unfair outcome', *Journal of Applied Social Psychology*, vol. 34(2), pp. 322–41.